JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1969, 2, 3-13 NUMBER 1 (SPRING 1969)

A TOKEN REINFORCEMENT PROGRAM IN A PUBLIC SCHOOL: A REPLICATION AND SYSTEMATIC ANALYSIS[1]

K. D. O'LEARY, W. C. BECKER, M. B. EVANS, AND R. A. SAUDARGAS

STATE UNIVERSITY OF NEW YORK AT STONY BROOK, UNIVERSITY OF ILLINOIS, AND FLORIDA STATE UNIVERSITY

A base rate of disruptive behavior was obtained for seven children in a second-grade class of 21 children. Rules, Educational Structure, and Praising Appropriate Behavior while Ignoring Disruptive Behavior were introduced successively; none of these procedures consistently reduced disruptive behavior. However, a combination of Rules, Educational Structure, and Praise and Ignoring nearly eliminated disruptive behavior of one child. When the Token Reinforcement Program was introduced, the frequency of disruptive behavior declined in five of the six remaining children. Withdrawal of the Token Reinforcement Program increased disruptive behavior in these five children, and reinstatement of the Token Reinforcement Program reduced disruptive behavior in four of these five. Follow-up data indicated that the teacher was able to transfer control from the token and back-up reinforcers to the reinforcers existing within the educational setting, such as stars and occasional pieces of candy. Improvements in academic achievement during the year may have been related to the Token Program, and attendance records appeared to be enhanced during the Token phases. The Token Program was utilized only in the afternoon, and the data did not indicate any generalization of appropriate behavior from the afternoon to the morning.

Praise and other social stimuli connected with the teacher's behavior have been established as effective controllers of children's behavior (Allen, Hart, Buell, Harris, and Wolf, 1964; Becker, Madsen, Arnold, and Thomas, 1967; Brown and Elliot, 1965; Hall, Lund, and Jackson, 1968; Harris, Johnston, Kelley, and Wolf, 1964; Harris, Wolf, and Baer, 1964; Scott, Burton, and Yarrow, 1967; Zimmerman and Zimmerman, 1962). When the teacher's use of praise and social censure is not effective, token reinforcement programs are often successful in controlling children (Birnbrauer, Wolf, Kidder, and Tague, 1965; Kuypers, Becker, and O'Leary, 1968; O'Leary and Becker, 1967; Quay, Werry, McQueen, and Sprague, 1966; Wolf, Giles, and Hall, 1968).

The token reinforcement program utilized by O'Leary and Becker (1967) in a third-grade adjustment class dramatically reduced disruptive behavior. In order to maximize the possibility of reducing the disruptive behavior of the children, O'Leary and Becker used several major variables simultaneously. The first objective of the present study was to analyze the separate effects of some of the variables utilized in the former study. More specifically, the aim was to examine the separate effects of Classroom Rules, Educational Structure, Teacher Praise, and a Token Reinforcement Program on children's disruptive behavior. Rules consisted of a list of appropriate behaviors that were reviewed daily. Educational Structure was the organization of an academic program into specified 30-min lessons such as spelling and arithmetic. The second objective was to assess whether a Token Reinforcement Program used only in the afternoon had any

[1]Portions of this paper were presented to the American Psychological Association, September, 1968, San Francisco, California. This research was supported primarily by Research Grant HD 00881-05 to Wesley C. Becker from the National Institutes of Health and secondarily by a Biomedical Science Grant 31-8200 to K. Daniel O'Leary from the State University of New York at Stony Brook. The authors are grateful to Nancy Brown, Connie Dockterman, Pearl Dorfmann, Jeanne Kappauf, Margery Lewy, Stanley Madsen, and Darlene Zientarski who were the major observers in this study. Appreciation for support of this study is expressed to Dr. Lowell Johnson, Director of Instruction, Urbana Public Schools, and to Mr. Richard Sturgeon, elementary school principal. The greatest thanks goes to Mrs. Linda Alsberg, the teacher who executed the Token Reinforcement Program and tolerated the presence of observers both morning and afternoon for eight months. Her patience and self-control during the Praise and Withdrawal Phases of the program were especially appreciated. Reprints may be obtained from K. Daniel O'Leary, Dept. of Psychology, State University of New York at Stony Brook, Stony Brook, N.Y. 11790.

effect on the children's behavior in the morning. Third, the present study sought to examine the extent to which the effects of the Token Reinforcement Program persisted when the Token Program was discontinued.

METHOD

Subjects

Seven members of a second-grade class of 21 children from lower-middle class homes served. At the beginning of the school year, the class had a mean age of 7 yr, 5 months, a mean IQ score of 95 (range 80 to 115) on the California Test of Mental Maturity, and a mean grade level of 1.5 on the California Achievement Test. The class was very heterogeneous with regard to social behaviors. According to the teacher, three of the children were quite well behaved but at least eight exhibited a great deal of undesirable behavior. The teacher, Mrs. A., had a master's degree in counseling but had only student teaching experience. She was invited to participate in a research project involving her class and received four graduate credits for participating in the project.

Observation

Children. Mrs. A. selected seven children for observation. All seven children were observed in the afternoon and four of the seven (S1, S2, S4, and S6) were also observed in the morning. Morning observations were made by a regular observer and a reliability checker from 9:30 to 11:30 every Monday, Wednesday, and Friday. Afternoon observations were made by two regular observers and a reliability checker from 12:30 to 2:30 every Monday, Wednesday, and Friday. Observations were made by undergraduate students who were instructed never to talk to the children or to make any differential responses to them in order to minimize the effect of the observers on the children's behavior. Before Base Period data were collected, the undergraduates were trained to observe the children over a three-week period in the classroom, and attention-seeking behaviors of the children directed at the observers were effectively eliminated before the Base Period.

Each child was observed for 20 min each day. The observers watched the children in a random order. Observations were made on a 20-sec observe, 10-sec record basis; *i.e.*, the observer would watch the child for 20 sec and then take 10 sec to record the disruptive behaviors which had occurred during that 20-sec period. The categories of behavior selected for observation were identical to those used by O'Leary and Becker (1967). Briefly, the seven general categories of disruptive behavior were as follows: (1) *motor behaviors:* wandering around the room; (2) *aggressive behaviors:* hitting, kicking, striking another child with an object; (3) *disturbing another's property:* grabbing another's book, tearing up another's paper; (4) *disruptive noise:* clapping, stamping feet; (5) *turning around:* turning to the person behind or looking to the rear of the room when Mrs. A. was in the front of the class; (6) *verbalization:* talking to others when not permitted by teacher, blurting out answers, name-calling; and (7) *inappropriate tasks:* doing arithmetic during the spelling lesson.

The present study was a systematic replication of O'Leary and Becker (1967). To facilitate comparison of the two studies, the dependent measure reported is the percentage of intervals in which one or more disruptive behaviors was recorded. Percentages rather than frequencies were used because the length of the observations varied due to unavoidable circumstances such as assemblies and snow storms. Nonetheless, most observations lasted the full 20 min, and no observation lasting less than 15 min was included.

Teacher. In order to estimate the degree to which the teacher followed the experimental instructions, Mrs. A. was observed by two undergraduates for 90 min on Tuesday and Thursday afternoons. Teacher behavior was not observed on Monday, Wednesday, and Friday when the children were observed because Mrs. A. understandably did not wish to have as many as five observers in the room at one time. Furthermore, because Mrs. A. was somewhat reluctant to have three regular observers and one or two graduate students in the room at most times, she was informed of the need for this observational intrusion and the mechanics thereof. This explanation made it impossible to assess the teacher's behavior without her knowledge, but it was felt that deception about teacher observation could have been harmful both to this project and future projects in the school. Nonetheless, fre-

quent teacher observations by two graduate students who were often in the room the entire week ensured some uniformity of her behavior throughout the week. The graduate students frequently met with Mrs. A. to alert her to any deviations from the experimental instructions, and equally important, to reinforce her "appropriate" behavior. Observations of the teacher's behavior were made on a 20-sec observe, 10-sec record basis. The categories of teacher behavior selected for observation were as follows:

I. Comments *preceding* responses.
   A. *Academic instruction:* "Now we will do arithmetic"; "Put everything in your desk"; "Sound out the words."
   B. *Social instruction:* "I'd like you to say 'please' and 'thank you'"; "Let me see a quiet hand"; "Let's sit up."

II. Comments *following* responses.
   A. *Praise:* "Good"; "Fine"; "You're right"; I like the way I have your attention."
   B. *Criticism:* "Don't do that"; "Be quiet"; "Sit in your seat!"
   C. *Threats:* "If you're not quiet by the time I count three . . . . "; "If you don't get to work you will stay after school"; "Do you want to stay in this group?"

The teacher's praise, criticism, and threats to individual children were differentiated from praise, criticism, and threats to the class as a whole. For example, "Johnny, be quiet!" was differentiated from "Class, be quiet!". Thus, eight different classes of teacher behavior were recorded: two classes of comments preceding responses and six classes following responses.

Procedure

The eight phases of the study were as follows: (1) Base Period, (2) Classroom Rules, (3) Educational Structure, (4) Praising Appropriate Behavior and Ignoring Disruptive Behavior, (5) Tokens and Back-up Reinforcement, (6) Praising Appropriate Behavior and Ignoring Disruptive Behavior (Withdrawal), (7) Tokens and Back-up Reinforcement, and (8) Follow-up. Three procedures, Educational Structure and both of the Token Reinforcement Phases, were instituted for a 2-hr period during the afternoon. The remainder of the procedures were in effect for the entire day. The eight procedures were in effect for all 21 children. The first four conditions were instituted in the order of hypothesized increasing effectiveness. For example, it was thought that Rules would have less effect on the children's behavior than the use of Praise. In addition, it was thought that the combination of Rules and Praise would have less effect than the Tokens and Back-up Reinforcers.

Base Period. After the initial three-week observer training period, the children were observed on eight days over a six-week Base Period to estimate the frequency of disruptive pupil behavior under usual classroom conditions.[2] The teacher was asked to handle the children in whatever way she felt appropriate. During the Base Period, Mrs. A. instructed all the children in subjects like science and arithmetic or took several students to small reading groups in the back of the room while the rest of the class engaged in independent work at their seats. Neither the particular type of activity nor the duration was the same each day. Stars and various forms of peer pressure were sporadically used as classroom control techniques, but they usually had little effect and were discontinued until experimentally reintroduced during the Follow-up Phase.

Classroom Rules. There were seven observations over a three-week period during the second phase of the study. The following rules or instructions were placed on the blackboard by the teacher: "We sit in our seats; we raise our hands to talk; we do not talk out of turn; we keep our desks clear; we face the front of the room; we will work very hard; we do not talk in the hall; we do not run; and, we do not disturb reading groups." Mrs. A. was asked to review the rules at least once every morning and afternoon, and frequent observations and discussions with Mrs. A. guaranteed that this was done on most occasions. The classroom activities again consisted of reading groups and independent seat work.

[2]Ten of the 18 observations during the Base Period were eliminated because movies were shown on those days, and disruptive behavior on those days was significantly less than on days when movies were not shown. Although movies were seldom used after Base Period, the seven subsequent observations when movies occurred were eliminated.

Educational Structure. It has been stated that a great deal of the success in token reinforcement programs may be a function of the highly structured regimen of the program and not a function of reinforcement contingencies. Since the Token Phase of the program was designed to be used during structured activities that the teacher directed, Mrs. A. was asked to reorganize her program into four 30-min sessions in the afternoon in which the whole class participated, *e.g.*, spelling, reading, arithmetic, and science. Thus, the purpose of the Educational Structure Phase was to assess the importance of structure *per se*. Mrs. A. continued to review the rules twice a day during this phase and all succeeding phases. During this phase there were five observations over a two-week period.

Praise and Ignore. In addition to Rules and Educational Structure, Mrs. A. was asked to praise appropriate behavior and to ignore disruptive behavior as much as possible. For example, she was asked to ignore children who did not raise their hands before answering questions and to praise children who raised their hands before speaking. In addition, she was asked to discontinue her use of threats. During this phase there were five observations over a two-week period.

Token I. Classroom Rules, Educational Structure, and Praise and Ignoring remained in effect. The experimenter told the children that they would receive points or ratings four times each afternoon. The points which the children received on these four occasions ranged from 1 to 10, and the children were told that the points would reflect the extent to which they followed the rules placed on the blackboard by Mrs. A. Where possible, these points also reflected the quality of the children's participation in class discussion and the accuracy of their arithmetic or spelling. The children's behavior in the morning did not influence their ratings in the afternoon. If a child was absent, he received no points. The points or tokens were placed in small booklets on each child's desk. The points were exchangeable for back-up reinforcers such as candy, pennants, dolls, comics, barrettes, and toy trucks, ranging in value from 2 to 30 cents. The variety of prizes made it likely that at least one of the items would be a reinforcer for each child. The prizes were on display every afternoon, and the teacher asked each child to select the prize he wished to earn before the rating period started.

During the initial four days, the children were eligible for prizes just after their fourth rating at approximately 2:30. Thereafter, all prizes were distributed at the end of the day. For the first 10 school days the children could receive prizes each day. There were always two levels of prizes. During the first 10 days, a child had to receive at least 25 points to receive a 2 to 5¢ prize (level one prize) or 35 points to receive a 10¢ prize (level two prize). For the next six days, points were accumulated for two days and exchanged at the end of the second day. When children saved their points for two days, a child had to receive 55 points to receive a 10¢ prize or 70 points to receive a 20¢ prize. Then, a six-day period occurred in which points were accumulated for three days and exchanged at the end of the third day. During this period, a child had to receive 85 points to receive a 20¢ prize or 105 points to receive a 30¢ prize. Whenever the prizes were distributed, the children relinquished all their points. During Token I, there were 13 observations over a five-week period.

For the first week, the experimenter repeated the instructions to the class at the beginning of each afternoon session. Both the experimenter and Mrs. A. rated the children each day for the first week in order to teach Mrs. A. how to rate the children. The experimenter sat in the back of the room and handed his ratings to Mrs. A. in a surreptitious manner after each rating period. Mrs. A. utilized both ratings in arriving at a final rating which she put in the children's booklets at the end of each lesson period. The method of arriving at a number or rating to be placed in the child's booklet was to be based on the child's improvement in behavior. That is, if a child showed any daily improvement he could receive a rating of approximately 5 to 7 so that he could usually earn at least a small prize. Marked improvement in behavior or repeated displays of relatively good behavior usually warranted ratings from 8 to 10. Ratings from 1 to 5 were given when a child was disruptive and did not evidence any daily improvement. Although such a rating system involves much subjective judgment on the part of the teacher, it is relatively easy to implement, and a subsidiary aim of the study was to assess whether a token system could be implemented by one

teacher in a class of average size. After the first week, the teacher administered the Token Program herself, and the experimenter was never present when the children were being observed. If the experimenter had been present during the Token Phases but not during Withdrawal, any effects of the Token Program would have been confounded by the experimenter's presence.

Withdrawal. To demonstrate that the token and back-up reinforcers and not other factors, such as the changes that ordinarily occur during the school year, accounted for the observed reduction in disruptive behavior, the token and back-up reinforcers were withdrawn during this phase. There were seven observations over a five-week period. When the prizes and the booklets were removed from the room, Mrs. A. told the children that she still hoped that they would behave as well as they had during the Token Period and emphasized how happy she was with their recent improvement. Rules, Educational Structure, and Praise and Ignoring remained in effect.

Token II. When the tokens and back-up reinforcers were reinstated, the children obtained a prize on the first day if they received 25 to 35 points. For the next four days there was a one-day delay between token and back-up reinforcement; the remainder of the Token Reinstatement Period involved a two-day delay of reinforcement. The prize and point system was identical to that during Token I. During this phase, there were five observations over a two-week period.

Follow-up. The token and back-up reinforcers were again withdrawn in order to see if the appropriate behavior could be maintained under more normal classroom conditions. In addition to the continued use of Praise, Rules, and Educational Structure, it was suggested that Mrs. A. initiate the use of a systematic star system. Children could receive from one to three stars for good behavior twice during the morning and once during the afternoon. In addition, the children received extra stars for better behavior during the morning restroom break and for displaying appropriate behavior upon entering the room at 9:15 and 12:30. At times, extra stars were given to the best behaved row of children. The children counted their stars at the end of the day; if they had 10 or more stars, they received a gold star that was placed on a permanent wall chart. If a child received 7 to 9 stars, he received a green star that was placed on the chart. The boys' gold stars and the girls' gold stars were counted each day; and each member of the group with the greater number of gold stars at the end of the week received a piece of candy. In addition, any child who received an entire week of gold stars received a piece of candy. All children began the day without stars so that, with the exception of the stars placed on the wall chart, everyone entered the program at the same level.

Such a procedure was a form of a token reinforcement program, but there were important procedural differences between the experimental phases designated Token and Follow-up. The back-up reinforcers used during the Token Phases were more expensive than the two pieces of candy a child could earn each week during the Follow-up Phase. In addition, four daily ratings occurred at half-hour intervals in the afternoons during the Token Phases but not during Follow-up. On the other hand, stars, peer pressure, and a very small amount of candy were used in the Follow-up Phase. As mentioned previously, both stars and peer pressure had been used sporadically in the Base Period with little effect. Most importantly, it was felt that the procedures used in the Follow-up Phase could be implemented by any teacher. During this phase there were six observations over a four-week period.

Reliability of Observations

The reliabilities of child observations were calculated according to the following procedure: an agreement was scored if both observers recorded one or more disruptive behaviors within the same 20-sec interval; a disagreement was scored if one observer recorded a disruptive behavior and the other observer recorded none. The reliability of the measure of disruptive behavior was calculated for each child each day by dividing the number of intervals in which there was agreement that one or more disruptive behaviors occurred by the total number of agreements plus disagreements. An agreement was scored if both observers recorded the same behavior within the same 20-sec interval. A disagreement was scored if one observer recorded the behavior and the other did not. The reliability of a particular class of teacher behavior on any

one day was calculated by dividing the total number of agreements for that class of behaviors by the total number of agreements plus disagreements for that class of behaviors. Reliabilities were calculated differently for child behaviors and teacher behaviors because different types of dependent measures were utilized for children and the teacher, and it was felt that reliability measures should be reported for the specific dependent measures used.

At least one reliability check was made during the afternoon on every child during the Base Period, and one child had three.[3] The average reliability of the measure of disruptive behavior during the afternoons of the Base Period for each of the seven children ranged from 88 to 100%. The following figures represent the number of reliability checks and the average of those reliability checks after the Base Period through the first Token Period for each child: S1: 6, 86%; S2: 7, 94%; S3: 6, 94%; S4: 6, 93%; S5: 6, 87%; S6: 6, 84%; S7: 6, 97%. Because of the repeated high reliabilities, reliability checks were discontinued when the token and back-up reinforcers were reinstated; *i.e.*, no reliability checks were made during or after the Withdrawal Phase.

Adequate morning reliabilities were not obtained until the Rules Phase of the study. The following figures represent the number of reliability checks and the average of those reliability checks during the Rules Phase: S1: 3, 93%; S2: 4, 68%; S4: 3, 91%; S6: 3, 88%. Morning reliability checks after the Rules Phase were made approximately every three observations (approximately seven occasions) through the first Token Period. Average reliabilities of the four children during the Rules, Educational Structure, Praise and Ignore, and Token I Phases ranged from 92 to 99%.

Eleven reliability checks for the various classes of teacher behavior before the Praise and Ignore Phase was introduced yielded average reliabilities as follows: academic instruction, 75%; social instruction, 77%; praise to individuals, 77%; praise to the class, 94%; criticism to individuals, 73%; criticism to the class, 72%; threats to individuals, 83%; and threats to the class, 83%.

RESULTS

Child Behavior

Figures 1 and 2 present morning and afternoon data; some of the variability within conditions can be seen. Figure 3 presents data of individual children as well as an average of seven children across afternoon conditions. An analysis of variance was performed on the percentages of combined disruptive behavior, averaged within the eight afternoon experimental conditions, for the seven subjects (See Fig. 3). The analysis of variance for repeated measures (Winer, 1962, p. 111) indicated differences among the eight experimental conditions ($F = 7.3$; $df = 7, 42$; $p < 0.001$). On the other hand, the percentages of combined disruptive behavior of the four children observed in the morning, averaged within conditions, did not change during Rules, Educational Structure, Praise and Ignore, or Token I ($F = 1.0$; $df = 4, 12$). Differences among afternoon conditions were assessed by *t*-tests. Significant and nonsignificant differences are grouped individually in Table 1.[4]

It should be emphasized that comparisons between Follow-up and Praise and Ignore are more meaningful than comparisons between Follow-up and Base, Rules, or Educational Structure. Praise and Follow-up were similar procedures; both included Rules, Educational Structure, and Praise and Ignore. The Base Period did not include any of these. Furthermore, after Rules and Educational Structure were initiated, Mrs. A. stated that she required more academic work from the children than during Base Period. A statistical analysis of the group data suggests that a token reinforcement program can reduce disruptive behavior and that a token reinforcement program can be replaced with a variant of a token program without an increase in disruptive behavior. However, a more detailed analysis of the data for individual children indicated that the Token Reinforcement Program was more effective for some children than others.

[3]Before 10 of the 18 observation days during the Base Period were eliminated because movies were shown on those days, at least three reliability checks had been made during the afternoon on each child.

[4]Two-tailed tests.

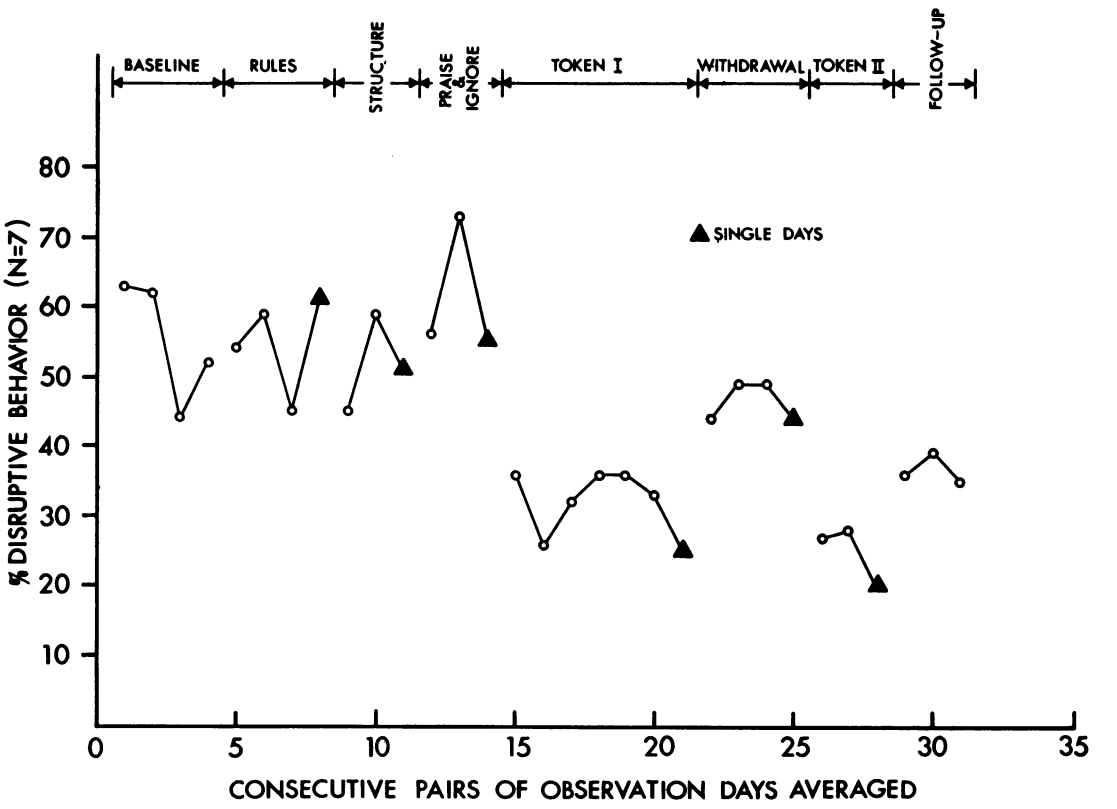

Fig. 1. Average percentage of combined disruptive behavior of seven children during the afternoon over the eight conditions: Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

The introduction of Rules, Educational Structure, and Praise and Ignore did not have any consistent effects on behavior (see Fig. 3). Praising Appropriate Behavior and Ignoring Disruptive Behavior deserve special mention. Although Mrs. A. used criticism occasionally during the Praise and Ignore Phase, she generally ignored disruptive behavior and used praise frequently. Initially, a number of children responded well to Mrs. A.'s praise, but two boys (S2 and S4) who had been disruptive all year became progressively more unruly during the Praise and Ignore Phase. Other children appeared to observe these boys being disruptive, with little or no aversive consequences, and soon became disruptive themselves. Relay races and hiding under a table contributed to the pandemonium. Several children were so disruptive that the academic pursuits of the rest of the class became impossible. The situation became intolerable, and the Praise and Ignore Phase had to be discontinued much earlier than had been planned.

The disruptive behavior of S7 was reduced to a very low level of 15% by a combination of Rules, Educational Structure, and Praise and Ignore. In the previous token program (O'Leary and Becker, 1967), in which a number of variables including rules, praise, educational structure, and a token program were simultaneously introduced, disruptive behavior during the token period was reduced to a level of 10%. Thus, the present Token Reinforcement Program probably would not be expected further to reduce disruptive behavior in this child.

During Token I, there was a marked reduction ($\geqq$ 18%) in the disruptive behavior of five children (S1, S2, S3, S4, and S6) and a reduction of 3% in S5. Withdrawal of the Token Program increased disruptive behavior from 5% to 45% in these six children. Reinstatement of the Token Program led to a decrease in five of these six children (S1, S2, S3, S4, S5). The disruptive behavior of five children (S1, S2, S4, S5, and S6) ranged from 8% to 39% lower during the Follow-up than dur-

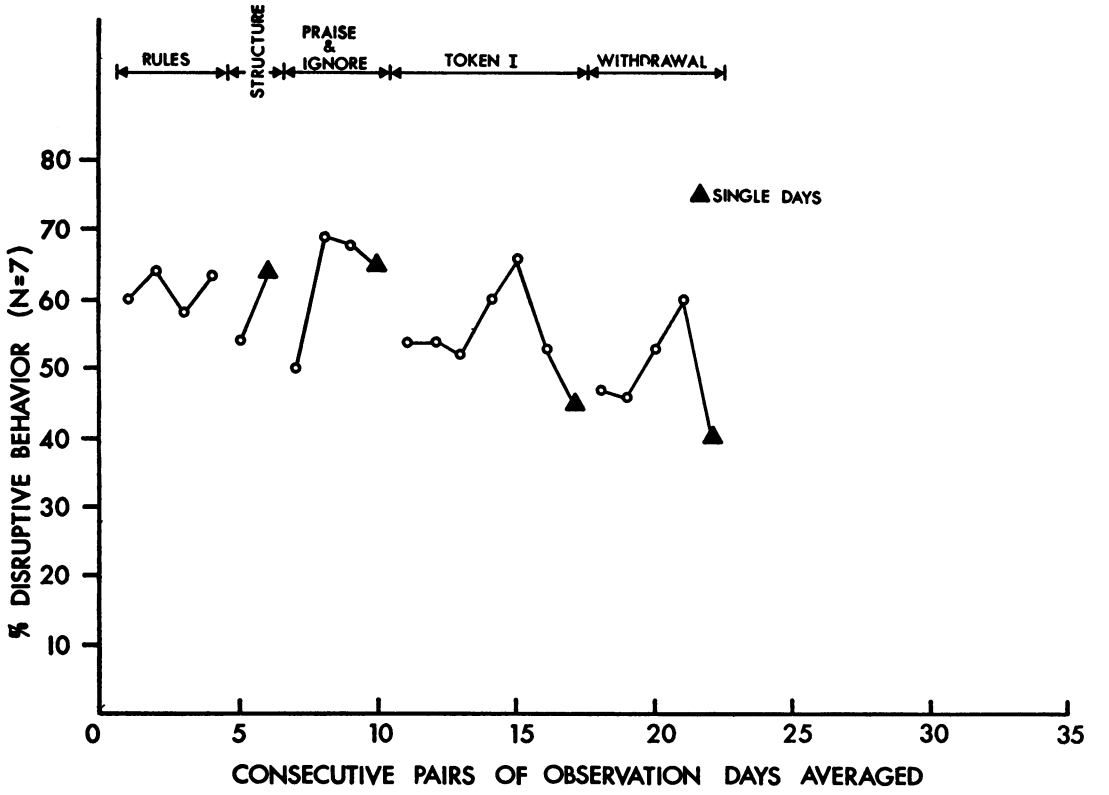

Fig. 2. Average percentage of combined disruptive behavior of four children during the morning over five conditions: Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

ing the Praise and Ignore Phase of the study. Since on no occasion did the Follow-up procedures precede Token I and/or Token II, this study did not demonstrate that Token I and/or Token II were necessary conditions for the success of the Follow-up procedures.

In summary, Token I and Token II were definitely associated with a reduction of disruptive behavior, *and* the Follow-up procedure was effective with three of the six children (S1, S2, and S4) who had more than 15% disruptive behavior during the Praise and Ignore Phase (S7 had 15% disruptive behavior during the Praise and Ignore Phase). Token I and Token II were associated with marked reductions of disruptive behavior of S3, but the frequency of disruptive behavior during the Follow-up was not substantially lower than during the Praise and Ignore Phase. Definitive conclusions concerning the effects of the Token Program cannot be drawn for S5 and S6, although some reduction of disruptive behavior was associated with either Token I and Token II for both of these children. In

Table 1

Significant		*Non-Significant*	
Token I *vs.* Withdrawal	$t = 3.3^{**}$	Rules *vs.* Educational Structure	$t = 0.8$
Token II *vs.* Withdrawal	$t = 2.9^{*}$	Educational Structure *vs.* Praise	$t = 1.0$
Token I *vs.* Praise	$t = 3.4^{**}$	Base *vs.* Withdrawal	$t = 1.2$
Token II *vs.* Praise	$t = 3.0^{*}$	Token I *vs.* Follow-up	$t = 1.1$
Base *vs.* Follow-up	$t = 3.2^{**}$	Token II *vs.* Follow-up	$t = 1.5$
Praise *vs.* Follow-up	$t = 3.3^{**}$		
Withdrawal *vs.* Follow-up	$t = 3.2^{**}$		

$^{**}p < 0.02$, $df = 6$
$^{*}p < 0.05$, $df = 6$

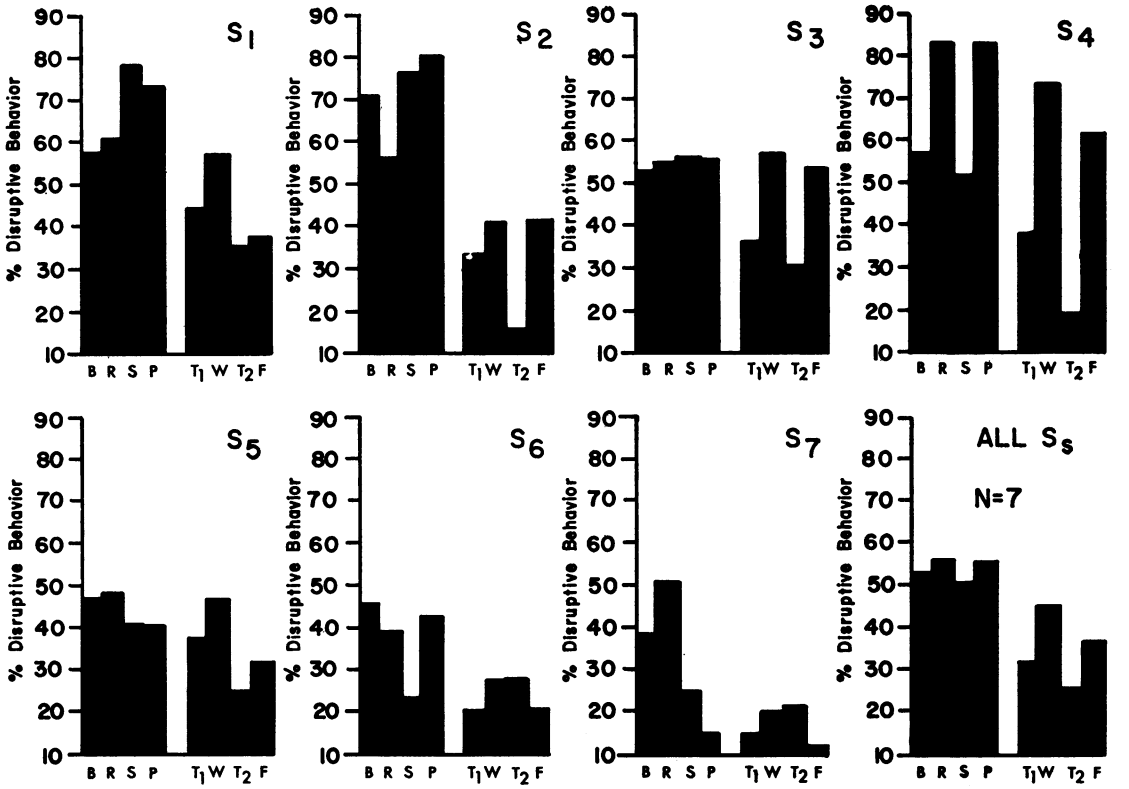

Fig. 3. Percentage of combined disruptive behavior for each of seven children during the eight conditions: Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

addition, the disruptive behavior of S5 and S6 was 8% and 20% less respectively during Follow-up than during the Praise and Ignore Phase.

Teacher Behavior

On any one day, the percentage of each of the eight classes of teacher behavior was calculated by dividing the number of intervals in which a particular class of behavior occurred by the total number of intervals observed on that day. Percentages rather than frequencies were used because of slight variations from the usual 90-min time base.

The percentages of different classes of teacher behavior were averaged within two major conditions: (1) data before Praise and Ignore Phase, and (2) data in the Praise and Ignore and succeeding Phases. The data in Fig. 4 show that in the Praise and Ignore Phase, Mrs. A. increased use of praise to individual children from 12% to 31% and decreased use of criticism to individuals from 22% to 10%. Mrs. A. also increased use of praise to the class from 1% to 7% and decreased criticism directed to the class from 11% to 3%. Because the frequency of threats was quite low, threats to individuals and threats to the class were combined in one measure. Using this combined measure, Mrs. A.'s use of threats decreased from 5% to 1%. There were no differences in Mrs. A.'s use of academic or social instruction. Consequently, the changes in the children's disruptive behavior can probably be attributed to contingencies and not to Mrs. A.'s use of cues concerning the desired behaviors.

DISCUSSION

Although a Token Reinforcement Program was a significant variable in reducing disruptive behavior in the present study, the results are less dramatic than those obtained by O'Leary and Becker, (1967). A number of factors probably contributed to the difference in effectiveness of the programs. The average of disruptive behavior during the Base Period in the 1967 study was 76%; in the present study it was 53%. The gradual in-

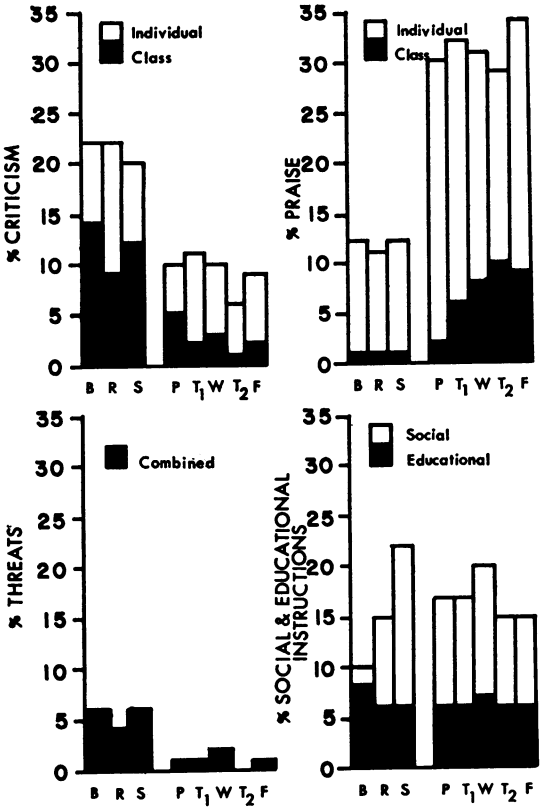

Fig. 4. Percentage of various teacher behaviors to individuals and to the class during the eight conditions: Base, Rules, Educational Structure, Praise and Ignore, Token I, Withdrawal, Token II, Follow-up.

troduction of the various phases of the program was probably less effective than a simultaneous introduction of all the procedures, as in the previous study. In the earlier study, the children received more frequent ratings. Five ratings were made each day at the introduction of the 1.5-hr token program, and they were gradually reduced to three ratings per day. In the present study, the children received four ratings per day during a 2-hr period. In the 1967 study, the class could earn points for popsicles by being quiet while the teacher placed ratings in the children's booklets; in the present study, group points were not incorporated into the general reinforcement program. In the 1967 study, the teacher attended a weekly psychology seminar where teachers discussed various applications of learning principles to classroom management. An *esprit de corps* was generated from that seminar that probably increased the teacher's commitment to change the children's behavior. Although Mrs. A. received graduate credits for her extensive participation in the project, she did not attend a seminar in classroom management. A number of children in the present study had an abundance of toys at home and it was difficult to obtain inexpensive prizes which would serve as reinforcers; in the earlier study, selection of reinforcers was not a difficult problem, since the children were from disadvantaged homes.

Related Gains

Academic. The 14 children for whom there were both pre- and post-measures on the California Achievement Test (including S1, S4, S5, S6, and S7) gained an average of 1.5 yr from October to June. The mean CAT score in October was 1.5 while the mean score in June was 3.0. Although there was no matched control group, such gains are greater than those usually obtained (Tiegs and Clark, 1963). While such gains are promising, conclusions about the effects of a token system on academic performance must await a more systematic analysis.

Attendance. Comparisons of the attendance records of the seven children during the observational days of the token and non-token phases yielded the following results: the average attendance percentage during the 45 observation days of Base, Rules, Educational Structure, Praise and Ignore, and Withdrawal was 86%. The average attendance percentage during the 20 observation days of Token I and Token II was 98%; the average attendance percentage during the 26 observation days of Token I, Token II, and Follow-up (a variant of a token program) was 99%. These attendance records are very encouraging, but because of the usual seasonal variations in attendance and the small sample of children, more definitive evidence is needed before conclusions about the effects of a token program on attendance can be made.

Cost of Program

The cost of the reinforcers in the present study was approximately $125.00. It is estimated that 3 hr of consulting time per week would be essential to operate a token reinforcement program effectively for one class in a public school. The cost of such a program and the amount of consulting time seem relatively small when compared to the hours

psychologists spend in therapy with children, often without producing significant behavioral changes (Levitt, 1963). Furthermore, as evidenced in the present study, control of behavior may be shifted from reinforcers, such as toys, to reinforcers existing within the natural educational setting, such as stars and peer prestige.

Generalization

During the morning, the majority of the children were engaged in independent seat work, while four or five children were in a reading group with the teacher in the back of the room. Although there were rules and frequent instructions during the morning, there was little reinforcement for appropriate behavior, since Mrs. A. felt that it would be disruptive to the rest of the class to interrupt reading groups to praise children who were doing independent work at their seats. Ayllon and Azrin (1964) found that instructions without reinforcement had little effect on the behavior of mental patients. Similarly, Rules (instructions) without reinforcement did not influence the behavior of the children in this study.

Mrs. A. was instructed to praise appropriate behavior and ignore disruptive behavior in the morning as well as the afternoon. However, Mrs. A.'s criteria of appropriate behavior in the morning differed from her criteria in the afternoon. For example, in the morning she often answered questions when a child failed to raise his hand before speaking. In the afternoon, on the other hand, she generally ignored a child unless he raised his hand. In order to achieve "generalization" of appropriate behavior in a Token Program such as this one, the teacher's response to disruptive behavior must remain constant throughout the day. The percentage of disruptive behavior was reduced during the morning of the first few days of Token I, but the children presumably learned to discriminate that their appropriate behavior was reinforced only in the afternoon. The differences in the children's behavior between the morning and the afternoon help to stress the point that "generalization" is no magical process, but rather a behavioral change which must be engineered like any other change.

REFERENCES

Allen, K. Eileen; Hart, Betty M., Buell, Joan S., Harris, Florence R., and Wolf, M. M. Effects of social reinforcement on isolate behavior of a nursery school child. *Child Development*, 1964, **35**, 511-518.

Ayllon, T. and Azrin, N. H. Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 327-331.

Becker, W. C., Madsen, C. H., Arnold, Carole R., and Thomas, D. R. The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1967, **1** (3), 287-307.

Birnbrauer, J. S., Wolf, M. M., Kidder, J. D., and Tague, Celia. Classroom behavior of retarded pupils with token reinforcement. *Journal of Experimental Child Psychology*, 1965, **2**, 219-235.

Brown, P. and Elliot, R. Control of aggression in a nursery school class. *Journal of Experimental Child Psychology*, 1965, **2**, 103-107.

Hall, R. V., Lund, Diane, and Jackson, Deloris. Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1968, **1**, 1-12.

Harris, Florence R., Johnston, Margaret K., Kelley, C. Susan, and Wolf, M. M. Effects of positive social reinforcement on regressed crawling of a nursery school child. *Journal of Educational Psychology*, 1964, **55**, 35-41.

Harris, Florence R., Wolf, M. M., and Baer, D. M. Effects of social reinforcement on child behavior. *Young Children*, 1964, **20**, 8-17.

Kuypers, D. S., Becker, W. C., and O'Leary, K. D. How to make a token system fail. *Exceptional Children*, 1968, **35**, 101-109.

Levitt, E. E. Psychotherapy with children: A further evaluation. *Behaviour Research and Therapy*, 1963, **1**, 45-51.

O'Leary, K. D. and Becker, W. C. Behavior modification of an adjustment class: A token reinforcement program. *Exceptional Children*, 1967, **33**, 637-642.

Quay, H. C., Werry, J. S., McQueen, Marjorie, and Sprague, R. L. Remediation of the conduct problem child in a special class setting. *Exceptional Children*, 1966, **32**, 509-515.

Scott, Phyllis M., Burton, R. V., and Yarrow, Marian R. Social reinforcement under natural conditions. *Child Development*, 1967, **38**, 53-63.

Tiegs, E. V. and Clark, W. W. Manual, California Achievement Tests, Complete Battery. 1963 Norms. California Test Bureau, Monterey, California.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

Wolf, M. M., Giles, D. K., and Hall R. V. Experiments with token reinforcement in a remedial classroom. *Behaviour Research and Therapy*, 1968, **6**, 51-64.

Zimmerman, Elaine H. and Zimmerman, J. The alteration of behavior in a special classroom situation. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 59-60.

Received 27 September 1968.